

### Mikrozensus Scientific Use File 2002: Dokumentation und Datenaufbereitung

Shahla, Hossein

Veröffentlichungsversion / Published Version  
Arbeitspapier / working paper

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:  
GESIS - Leibniz-Institut für Sozialwissenschaften

#### Empfohlene Zitierung / Suggested Citation:

Shahla, H. (2005). *Mikrozensus Scientific Use File 2002: Dokumentation und Datenaufbereitung*. (ZUMA-Methodenbericht, 2005/01). Mannheim: Zentrum für Umfragen, Methoden und Analysen -ZUMA-. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-48489-3>

#### Nutzungsbedingungen:

Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

#### Terms of use:

This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

*ZUMA-Methodenbericht 2005/01*

**Mikrozensus Scientific Use File 2002**  
**Dokumentation und Datenaufbereitung**

Hossein Shahla

Januar 2005

ISSN 1610-9953

ZUMA  
German Microdata Lab (GML)  
Quadrat B2,1  
Postfach 12 21 55  
68072 Mannheim

Telefon: (0621) 1246 - 253

Telefax: (0621) 1246 - 100

E-mail: [mikrodaten@zuma-mannheim.de](mailto:mikrodaten@zuma-mannheim.de)

# INHALTSVERZEICHNIS

<b>1. Einleitung .....</b>	<b>1</b>
<b>2. Mikrozensus 2002 .....</b>	<b>1</b>
2.1 Kurzbeschreibung .....	1
2.2 Datenerhebung .....	2
2.3 Die Stichprobe .....	2
2.4 Das Erhebungsprogramm .....	3
<b>3. Der faktisch anonymisierte Mikrozensus 2002 – Das Scientific Use File .....</b>	<b>4</b>
<b>4. Aufbereitung des Rohdatenfiles zum Mikrozensus Scientific Use File 2002 und     Überprüfung der Plausibilität des erstellten Systemfiles .....</b>	<b>5</b>
4.1 Einlesen des Rohdatenmaterials und Erstellung des Systemfiles .....	5
4.2 Gewichtung und Hochrechnung zum Vergleich mit den Veröffentlichungen des Statistischen Bundesamtes .....	6
4.3 Bevölkerungsbegriffe und Abgrenzungen im Mikrozensus .....	7
4.4 Plausibilitätsprüfung des aufbereiteten Mikrozensus Scientific Use Files 2002 .....	8
 <b>Anhang:</b>	
A) Randverteilung ausgewählter Variablen des Scientific Use Files 2002 .....	10
B) SPSS-Syntax für die Plausibilitätsprüfungen des Scientific Use Files 2002 .....	14
C) Weiterführende Literatur .....	16

## **1. Einleitung**

In dem vorliegenden ZUMA-Methodenbericht werden nach einer kurzen Beschreibung des Erhebungsprogramms, der Stichprobe und der Besonderheiten des Mikrozensus 2002 die einzelnen Schritte der Aufbereitung der faktisch anonymisierten Mikrozensusdaten (des Scientific Use Files) dargestellt.

Die Bereitstellung der faktisch anonymisierten Daten als Rohdatenfile in ASCII-Format erfolgt durch das Statistische Bundesamt. In Kooperation mit dem Statistischen Bundesamt stellt das German Microdata Lab dem Nutzer ausführliche Informationen über die Scientific Use Files zur Verfügung und bietet Setups zur Aufarbeitung des Datenmaterials und zur Erstellung von System-Files in Datenformaten gängiger Softwarepakete an. Ausführlichere Informationen über die Mikrozensus Scientific Use Files (vorhandene Variablen, verwendete Systematiken, Fragebogen, Schlüsselverzeichnis, intertemporale Vergleichbarkeit, Mikrodaten Tools etc.) werden auch in ZUMA-Methodenberichten zum intertemporalen Vergleich von Mikrozensen<sup>1</sup> und im World Wide Web des German Microdata Lab<sup>2</sup> bereitgestellt und können dort abgerufen werden.

Im Folgenden wird nach einer kurzen Beschreibung des Mikrozensus und des Mikrozensus Scientific Use Files 2002 der Vorgang der Datenaufbereitung beschrieben und die notwendigen Schritte zur Überprüfung der Plausibilität von eingelesenen Daten erläutert.

## **2. Mikrozensus 2002**

### **2.1 Kurzbeschreibung**

Der Mikrozensus ist mit einem Auswahlatz von 1% der Bevölkerung die größte jährliche Haushaltsbefragung in Europa. Sie wird seit 1957 in Westdeutschland und seit 1991 in den neuen Bundesländern durchgeführt. Grundlage für die ausführlichen Angaben zu bevölkerungs- und arbeitsmarktstatistischen Strukturdaten des Mikrozensus ist die gesamte wohnberechtigte Bevölkerung in Deutschland, zu der im April 2002 rund 82,5 Mio. Menschen gehörten. Der Mikrozensus ist durch eine hohe Kontinuität im Frageprogramm und aufgrund der Auskunftspflicht durch einen niedrigen Unit-Nonresponse gekennzeichnet. Der Mikrozensus

---

<sup>1</sup> Vgl. ZUMA-Methodenberichte 2005/02, 2004/09, 2003/08 und 2003/02 sowie die ZUMA-Technischen Berichte 2000/08, 2000/09 und 1998/09.

<sup>2</sup> Vgl.: [http://www.gesis.org/Dauerbeobachtung/Mikrodaten/Service\\_Guide/index.htm](http://www.gesis.org/Dauerbeobachtung/Mikrodaten/Service_Guide/index.htm)

ermöglicht differenzierte Analysen auf Personen-, Familien- und Haushaltsebene sowohl im Quer- als auch im Längsschnitt und eignet sich für die Beobachtung von Strukturveränderungen sowie zu tiefgegliederten Analysen.

## **2.2 Datenerhebung**

Die Angaben im Mikrozensus 2002 beziehen sich auf die Berichtswoche vom 22. bis 28. April 2002. Die Daten des Mikrozensus werden durch die Statistischen Landesämter erhoben. Die Befragungen finden i.d.R. mündlich statt, ein kleiner Teil erfolgt mittels schriftlicher Auskunft der Befragten. Für alle Beteiligten besteht gemäß der gesetzlichen Regelungen Auskunftspflicht. Nur wenige Fragen sind freiwillig zu beantworten. Der Mikrozensus ist eine Haushaltsstichprobe, d.h. alle Personen in einem Haushalt werden entweder direkt befragt, oder eine erwachsene Person gibt Auskunft über die weiteren (alle minderjährigen) Haushaltsmitglieder. Der Mikrozensus umfasst auch ausländische Haushalte. Zur Erhebung ihrer Angaben wird der gleiche Fragebogen verwendet wie bei deutschen Haushalten. Die Daten werden von den Statistischen Landesämtern aufbereitet und beim Statistischen Bundesamt zu einem Datensatz zusammengestellt.

## **2.3 Die Stichprobe**

Für den Mikrozensus 2002 gilt das Stichprobendesign von 1990. Die Stichprobe wurde aus dem Datenmaterial der Volkszählung 1987 gezogen. Das stichprobenmethodische Grundkonzept des Mikrozensus ist die einstufige Klumpenstichprobe. Die Klumpen sind Auswahlbezirke, die ganze Gebäude oder Gebäudeteile umfassen. Zur Verbesserung der Regionalisierbarkeit wurde 1990 die durchschnittliche Größe der Auswahlbezirke verkleinert. Ein Auswahlbezirk umfasst durchschnittlich 9 Wohnungen (vor 1990: 23 Wohnungen). Die Stichprobe wird jährlich durch eine Ergänzungsauswahl der Bautätigkeitsstatistik aktualisiert.

Während für das frühere Bundesgebiet die Grundausswahl auf der Basis der Volkszählung 1987 vorgenommen werden konnte, war für die neuen Bundesländer die Volkszählung 1981 als Auswahlgrundlage zu alt. Aus diesem Grund wurde in den neuen Bundesländern und Ost-Berlin das „Bevölkerungsregister Statistik“ in einem weitgehend analogen Verfahren zur Abgrenzung der Auswahlinheiten herangezogen. Im Mikrozensus 2002 sind Informationen zu dem Auswahlbezirk, der Gebäudeschicht, der Bogenart und der Zahl der Haushalte enthalten.

## 2.4 Das Erhebungsprogramm

Das Fragenprogramm des Mikrozensus besteht aus einem gleichbleibenden Grundprogramm, das sozioökonomische Grundinformationen liefert und die laufende Beobachtung des Arbeitsmarktes ermöglicht. Im Grundprogramm, das sich an alle Haushalte richtet, sind im Mikrozensus 2002 folgende Bereiche enthalten:

- Fragen zur Wohnung / zum Haushalt
- Fragen zu den Personen im Haushalt
- Schulbesuch
- Erwerbsbeteiligung
- Erste gegenwärtige Erwerbstätigkeit
- Zweite Erwerbstätigkeit
- Arbeitsuche / Arbeitsplatzwechsel
- Altersvorsorge
- Ausbildung
- Unterhalt, Einkommen

Neben den jährlich zu erhebenden Merkmalen des Grundprogramms (Auswahlsatz 1%) werden in jährlichen Ergänzungsprogrammen mit einem Auswahlsatz von bis zu 0,45% weitere Themen erfragt.<sup>3</sup>

Im Ergänzungsprogramm 2002 sind folgende zusätzliche Bereiche enthalten:

- Berufliche und allgemeine Aus- und Fortbildung
- Ergänzende Angaben zur Erwerbstätigkeit
- Frühere Erwerbstätigkeit
- Erwerbsbeteiligung ein Jahr vor der Erhebung

Darüber hinaus gibt es ein Zusatzprogramm mit Merkmalen, die nur im Abstand von 4 Jahren erhoben werden. Im Rahmen der vierjährigen Zusatzprogramme enthält der Mikrozensus 2002 folgende Angaben:

- Fragen zur Wohnsituation

---

<sup>3</sup> In Anlehnung an die Genauigkeitsanforderung für die Arbeitskräfteerhebung der EU, dass der einfache relative Standardfehler auf Regierungsbezirksebene für Merkmale, die 5% der Bevölkerung im erwerbsfähigen Alter betreffen, höchstens 8% betragen darf, ist der Auswahlsatz des Ergänzungsprogramms auf Regierungsbezirksebene unterschiedlich (0,4%, 0,6%, 0,8% oder 1%). Im Bundesdurchschnitt beträgt er rund 0,45%.

### 3. Der faktisch anonymisierte Mikrozensus 2002 – Das Scientific Use File

Bei den vorliegenden Daten des Mikrozensus Scientific Use File 2002 handelt es sich um eine faktisch anonymisierte 70-Prozent-Substichprobe des Mikrozensus 2002. Für die faktisch anonymisierte Substichprobe gilt, dass sie als systematische Zufallsauswahl aus dem Originalmaterial des Statistischen Bundesamtes gezogen wurde. Das Datenmaterial wurde dazu nach Bundesland, Regierungsbezirk, Gemeindegrößenklasse, Anzahl der Personen in Privathaushalten, Auswahlbezirksnummer und Nummer der Wohnung sortiert. Durch diese Anordnung ist gewährleistet, dass die Substichprobe hinsichtlich dieser Merkmale nur geringe zufallsbedingte Abweichungen aufweist. In diesem Prozess wurden Gemeinschaftsunterkünfte wie Wohnungen behandelt, das Merkmal „Zahl der Personen in Privathaushalten“ wurde nicht zur Ziehung der Substichprobe herangezogen. Bei Gemeinschaftsunterkünften erhielt jede Person eine eigene fortlaufende ("Wohnungs"-)Nummer. Zur Stichprobenziehung diente die letzte (Einer-)Stelle der Wohnungsnummer. Alle Wohnungen, deren Einerstelle einer von drei zufällig gezogenen einstelligen Zahlen entsprach, gelangten nicht in die Stichprobe (30% der Wohnungen).<sup>4</sup>

Weiterhin unterscheidet sich das Scientific Use File vom Originalfile des Mikrozensus dadurch, dass bestimmte Variablen im Scientific Use File, bedingt durch die Anonymisierung, in klassierter und vergrößerter Form verfügbar sind.

So sind im Scientific Use File die Regionalangaben Bundesland und eine vergrößerte Gemeindegrößenklasse, die nicht für die kleinen Bundesländer Bremen und Saarland ausgewiesen ist, enthalten. Mit Hilfe eines speziellen Codes der Variablen Gemeindegrößenklasse kann zwischen West- und Ost-Berlin (EF708=9) unterschieden werden. Auch bei den weiteren Variablen des Scientific Use Files werden je nach Bedarf Merkmalsvergrößerungen vorgenommen, so dass jede Ausprägung in den univariaten Verteilungen mindestens 5.000 Fälle in der Grundgesamtheit aufweist. Die Ausprägungen der Variable Nationalität werden so aggregiert, dass jede Nationalitätengruppe in der Grundgesamtheit wenigstens 50.000 Einwohner umfasst.

---

<sup>4</sup> Das beschriebene Auswahlverfahren gilt für die anonymisierten Files der Mikrozensen mit Fragen zur Wohnsituation. Bei allen anderen Mikrozensus Scientific Use Files werden anstelle der Wohnungsnummern die Haushaltsnummern zur Stichprobenziehung herangezogen werden.

## **4. Aufbereitung des Rohdatenfiles zum Mikrozensus Scientific Use File 2002 und Überprüfung der Plausibilität des erstellten Systemfiles**

### **4.1 Einlesen des Rohdatenmaterials und Erstellung des Systemfiles**

Die vom German Microdata Lab bereitgestellten Setups für das Mikrozensus Scientific Use File 2002<sup>5</sup> dienen zum Einlesen des Rohdatenmaterials und zum Erstellen von Systemfiles für die Programmpakete SAS, SPSS und STATA. Sie beinhalten Programmanweisungen zum Umwandeln von nicht-numerischen Variablen in numerische Variablen,<sup>6</sup> zur Definition von fehlenden Werten sowie zum Versehen der Variablen und ihrer Ausprägungen mit entsprechenden Labeln.

Bis zum Mikrozensus 1996 wurden bei der Aufbereitung der Rohdaten die nicht fortlaufenden Kategoriennummern (z.B. 1, 3, 4) in fortlaufende Nummern (z.B. 1, 2, 3) umgewandelt. Der Schlüssel 0 'ohne Angabe' wurde i.d.R. in 8 'ohne Angabe' umkodiert. Um die größtmögliche Nähe des Scientific Use Files zum Originalfile zu gewährleisten, wurden diese Umkodierungen im Mikrozensus 2002 wie auch in den vorhergehenden Mikrozensen 2001, 2000, 1999, 1998 und 1997 nicht vorgenommen. Die fehlenden Werte wurden möglichst in 0 umgewandelt. War dieser Wert schon vergeben, dann wurden die fehlenden Werte in 9 oder 8 (bzw. 99, 999, 98) umkodiert.

Um eine reibungsfreie Aufbereitung der Daten zu gewährleisten, empfiehlt es sich, die in den Setups für das jeweilige Programmpaket vorgesehenen Voreinstellungen zu berücksichtigen und je nach Leistung des verarbeitenden Systems die Setups schrittweise durchführen zu lassen. Sind die Daten eingelesen und ist das entsprechende Systemfile erstellt, können die Verteilungen ausgewählter Variablen des Scientific Use Files 2002 (vgl. Anhang A) zur Überprüfung des fehlerfreien Ablaufs der Datenaufbereitung herangezogen werden.

Als Referenz zur Plausibilitätsprüfung der Scientific Use Files dienen die in den Fachserien des Statistischen Bundesamtes veröffentlichten Ergebnisse des jeweiligen Mikrozensus. Für den Mikrozensus 2002 sind dies die Reihe 3 (Haushalte und Familien), die Reihe 4.1.1 (Stand und Entwicklung der Erwerbstätigkeit) und die Reihe 4.1.2 (Beruf, Ausbildung und Arbeitsbedingungen der Erwerbstätigen) der Fachserie 1.<sup>7</sup> Bei dem Vergleich der Ergebnisse des Scientific Use Files mit den entsprechenden Veröffentlichungen des Statistischen Bundesamtes ist darauf zu achten, dass die in diesen Reihen veröffentlichten Zahlen gewichtete, auf die

---

<sup>5</sup> Die SAS-, SPSS- und STATA-Setups können von den Webseiten des German Microdata Lab:

[http://www.gesis.org/Dauerbeobachtung/Mikrodaten/Daten/Abteilungsdaten/Mikrozensen/mz\\_daten\\_index.htm](http://www.gesis.org/Dauerbeobachtung/Mikrodaten/Daten/Abteilungsdaten/Mikrozensen/mz_daten_index.htm) heruntergeladen werden.

<sup>6</sup> Im Mikrozensus Scientific Use File 2002 ist lediglich die Variable AF643, die in EF643 umgewandelt wird, davon betroffen.



Gesamtbevölkerung hochgerechnete Ergebnisse des Mikrozensus sind. Weiterhin beziehen sich die Zahlen auf nach bevölkerungs- oder erwerbstatistischen Konzepten definierte (Teil-) Populationen, die in der jeweiligen Reihe der Fachserien des Statistischen Bundesamtes erläutert werden.<sup>8</sup>

Im Folgenden werden die Modalitäten der Gewichtung und Hochrechnung der Scientific Use Files sowie die für den Vergleich mit den Veröffentlichungen des Statistischen Bundesamtes relevanten Bevölkerungsbegriffe und Abgrenzungen erläutert. Die konkreten Schritte und Ergebnisse der Plausibilitätsprüfung der aufbereiteten Mikrozensus Scientific Use Files 2002 beim German Microdata Lab werden in Abschnitt 4.4 dargestellt.

#### **4.2 Gewichtung und Hochrechnung zum Vergleich mit den Veröffentlichungen des Statistischen Bundesamtes**

Im Scientific Use File liegen fünf Gewichte zur Kompensation der systematischen Ausfälle und des zufallsbedingten Stichprobenfehlers sowie zur Anpassung der Daten an die Bevölkerungsfortschreibung vor:<sup>9</sup>

- EF750 Personenbezogener Gewichtungsfaktor
- EF751 Haushalts-/Familien-Gewichtungsfaktor in der regionalen Anpassungsschicht
- EF755 Personenbezogener Gewichtungs- und Hochrechnungsfaktor für die disproportionale 0,45%-Unterstichprobe/Ergänzungsprogramm (der Auswahlatz dieser Unterstichprobe liegt auf Regierungsbezirksebene bei 0,4%, 0,6%, 0,8% oder 1% und beträgt im Bundesdurchschnitt 0,45%).
- EF756 Gewichtungs- und Hochrechnungsfaktor im Regierungsbezirk (Haushalts- und Familienfaktor für das Ergänzungsprogramm)
- EF761 Wohnungsfaktor

Sollen die Ergebnisse von relativen Häufigkeiten des Scientific Use Files mit Veröffentlichungen des Statistischen Bundesamtes zum Mikrozensus 2002 verglichen werden, muss das entsprechende Gewicht des Mikrozensus (EF750, EF751, EF755, EF756) verwendet werden. Um auf 1% der Bevölkerung, also auf den vollen Mikrozensus hochzurechnen, wird das Gewicht mit dem Kehrwert der Ziehungswahrscheinlichkeit für das Scientific Use File (100/70) multipliziert (hier in SPSS-Syntax am Beispiel personenbezogener Auswertungen):

```
COMPUTE XXXX=EF750 * 100/70.
```

```
WEIGHT BY XXXX.
```

---

<sup>7</sup> Vgl. die jeweilige Reihe der Fachserie 1, Statistisches Bundesamt, Wiesbaden 2003, Metzler-Poeschel, Stuttgart.

<sup>8</sup> Vgl. den „Textteil“ der jeweiligen Reihe der Fachserie 1, Statistisches Bundesamt, Wiesbaden 2003, Metzler-Poeschel, Stuttgart.

<sup>9</sup> Hierbei handelt es sich um die Gewichtungs- und Hochrechnungsfaktoren des Mikrozensus Originalfiles. Für die Scientific Use Files stehen keine spezifischen Gewichtungsfaktoren zur Verfügung.

Wie erwähnt, beziehen sich die in den Fachserien des Statistischen Bundesamtes veröffentlichten absoluten Zahlen auf die gewichteten, auf die Gesamtbevölkerung hochgerechneten Daten des Mikrozensus. Zur Herstellung der Vergleichbarkeit zwischen den absoluten Häufigkeiten des Scientific Use Files mit den Veröffentlichungen des Statistischen Bundesamtes müssen indes die Ergebnisse eigener Berechnungen noch zusätzlich mit 100 (Kehrwert der Ziehungswahrscheinlichkeit für den vollen Mikrozensus) multipliziert werden. Dies entspricht der Multiplikation des jeweiligen Gewichtungsfaktors mit 100/0,7.

### 4.3 Bevölkerungsbegriffe und Abgrenzungen im Mikrozensus

Zum Vergleich der Ergebnisse des Scientific Use Files (SUF) mit den vom Statistischen Bundesamt veröffentlichten Zahlen sind die jeweiligen Bevölkerungs- und Erwerbskonzepte zu berücksichtigen. Um solche Vergleiche zu erleichtern, gibt die folgende Tabelle einen Überblick über die Bevölkerungsbegriffe, deren Abgrenzung sowie die entsprechenden Gewichtungs- und Hochrechnungsfaktoren.

Oft ist eine Kombination der Bevölkerung am Ort der Hauptwohnung und der Bevölkerung in Privathaushalten sinnvoll, um die Bevölkerung am Ort der Nebenwohnung und in Anstalts- oder Gemeinschaftsunterkünften ausschließen zu können. In diesem Fall sind die Zahlen mit dem Haushalts- und Familienfaktor (EF751) hochzurechnen. In Veröffentlichungen des Statistischen Bundesamtes zum Mikrozensus wird diese Kombination jedoch nicht verwendet.

**Tabelle 1: Bevölkerungsbegriffe, Abgrenzungen und Gewichtungen im Mikrozensus SUF 2002**

Bevölkerungsbegriff Subpopulation	Filter zur Abgrenzung der Subpopulation	Gewichtungsfaktor
Wohnberechtigte Bevölkerung	-	EF750 Personen-Gewichtungsfaktor
Bevölkerung am Ort der Hauptwohnung	EF505=1 or EF505=2	EF750 Personen-Gewichtungsfaktor
Erwerbstätige	(EF505=1 or EF505=2) and EF504=1	EF750 Personen-Gewichtungsfaktor
Erwerbstätige (EU-Arbeitskräfteerhebung, 0,45% Substichprobe)	(EF505=1 or EF505=2) and EF504=1 and EF738=1	EF755 Gewichtungsfaktor f. disprop. 0,45% Unterstichprobe
Privathaushalte	EF506=1 and EF507=1	EF751 Haushalts-/Familien- Gewichtungsfaktor
Bevölkerung in Privathaushalten	EF506=1	EF751 Haushalts-/Familien- Gewichtungsfaktor
Bevölkerung am Familienwohnsitz	EF508=1	EF75 Haushalts-/Familien- Gewichtungsfaktor
Familien	EF509=1 and EF508=1	EF751 Haushalts-/Familien- Gewichtungsfaktor
Nichteheliche Lebensgemeinschaften (verschieden geschlechtlich)	EF506=1 and EF515=1 and EF630=1-4	EF75 Haushalts-/Familien- Gewichtungsfaktor

#### 4.4 Plausibilitätsprüfung des aufbereiteten Mikrozensus Scientific Use Files 2002

Die Plausibilitätsprüfung des Mikrozensus Scientific Use Files 2002 erfolgte im German Microdata Lab durch den Vergleich des aufbereiteten Datensatzes mit den vom Statistischen Bundesamt veröffentlichten Tabellen in den entsprechenden Reihen der Fachserie 1. Dabei wurde nach den vom Statistischen Bundesamt verwendeten Bevölkerungs- und Erwerbskonzepten gewichtet und selektiert. Tabelle 2 enthält die Liste der zur Plausibilitätsprüfung herangezogenen Variablen, Angaben zu notwendigen Gewichtungen und Abgrenzungen sowie die entsprechenden Seitenzahlen der jeweiligen Datenreihe des Statistischen Bundesamtes.

**Tabelle 2: Gewichtung, Abgrenzung und Vergleich mit den Fachserien des Statistischen Bundesamtes für ausgewählte Variablen des Mikrozensus Scientific Use File 2002**

<b>Fachserie 1 Reihe 3 (Haushalte und Familien)</b>		
Gewichtung: Haushaltsgewicht		
Bevölkerung in Privathaushalten: EF506=1		
EF1:	Land der BRD	(Tab. 2.28, S.119)
EF30:	Alter (klassiert)	(Tab. 1.1, S.39)
EF32:	Geschlecht	(Tab. 1.1, S.39)
EF35:	Familienstand	(Tab. 1.1, S.39)
EF521:	Zahl der Personen in Privathaushalten	(Tab. 1.1, S.39)
EF541:	Haushaltstyp	(Tab. 2.23, S.110)
EF542:	Familientyp in den Haushaltstypen	(Tab. 2.23, S.110)
EF44:	Staatsangehörigkeit	(Tab. 2.19, S.102)
<b>Fachserie 1 Reihe 4.1.1 (Stand und Entwicklung der Erwerbstätigkeit)</b>		
Gewicht: Personengewicht		
Bevölkerung am Ort der Hauptwohnung: EF505=1 oder EF505=2		
Erwerbstätige: EF505=1 oder EF505=2 und EF504=1		
Erwerbslose: EF505=1 oder EF505=2 und EF504=2 oder EF504=3		
EF127:	Stellung im Beruf	(Tab. 1.2, S.41)
EF129:	Wirtschaftszweige	(Tab. 3.4, S.63)
EF138:	Vollzeit/Teilzeittätigkeit	(Tab. 3.17, S.100)
EF139:	Grund für Teilzeittätigkeit	(Tab. 3.17, S.100)
EF213:	Grund der Arbeitsuche bei NEWT	(Tab. 4.5, S. 114)
EF216:	Tätigkeit als Selbständiger/Arbeitnehmer	(Tab. 4.4, S.112)
EF217:	Art der gesuchten Tätigkeit	(Tab. 4.4, S.112)
EF239:	Dauer der Arbeitsuche	(Tab. 4.3, S.110)
EF248:	1. Aktive Methode der Arbeitsuche (Arbeitnehmer)	(Tab. 4.4, S.112)
EF287:	Allgemeinbildender Abschluss	(Tab. 3.16.1, S.92)
EF289:	Beruflicher Ausbildungsabschluss	(Tab. 3.16.1, S.92)
<b>Fachserie 1 Reihe 4.1.2 (Beruf, Ausbildung und Arbeitsbedingungen der Erwerbstätigen)</b>		
Gewichtung: a) Personengewicht (EF750); b) Personengewicht für die 0.45%-Unterstichprobe (EF755)		
Bevölkerung am Ort der Hauptwohnung und Erwerbstätige: EF505=1 oder EF505=2 und EF504=1		
a): EF128:	Beruf	(Tab. 6, S.35ff)
b): EF147:	Samstagsarbeit	(Tab. 14, S. 98)
EF150:	Nachtschicht	(Tab. 16, S. 95)

Die Programmanweisungen für die durchgeführten Plausibilitätsprüfungen sind als SPSS-Syntax im Anhang B aufgeführt.

Bedingt durch die Ziehung der 70%-Stichprobe ergeben sich Abweichungen zwischen den Häufigkeiten der Variablen des Scientific Use Files und den in Fachserien des Statistischen Bundesamtes veröffentlichten Zahlen. Die Mehrzahl der oben aufgeführten Variablen weichen i.d.R. nur in geringem Maße (meist 0% bis 1%, max. 5%) von den veröffentlichten Daten ab. Abweichungen über 5% wurden bei Variablen festgestellt, deren Merkmalsausprägungen mit sehr kleinen Fallzahlen besetzt sind (z.B. Staatsangehörigkeit und Wirtschaftsbereiche).

**Anhang A: Randverteilung ausgewählter Variablen des Scientific Use Files 2002**

Bevölkerungskonzept	Bevölkerung in Privat- haushalten	Bevölkerung am Ort der Hauptwoh- nung	Bevölkerung am Familien- wohnsitz	Erwerbstätige Bevölkerung am Ort der Hauptwoh- nung	Fallzahlen
	<i>Haushalts- gewicht<sup>10</sup></i>	<i>Personenge- wicht<sup>11</sup></i>	<i>Haushaltsge- wicht</i>	<i>Personenge- wicht</i>	<i>ohne Selektion und ohne Ge- wichtung</i>
Merkmal	<i>select if (ef506=1)</i>	<i>select if (ef505=1 or ef505=2)</i>	<i>select if (ef508=1)</i>	<i>select if (ef505=1 or ef505=2) and (ef504=1)</i>	
Insgesamt	<b>827557</b>	<b>823840</b>	<b>816935</b>	<b>365073</b>	<b>503075</b>
<b>EF32: Geschlecht</b>					
Männlich	402665	402421	396921	203396	242660
Weiblich	424891	421419	420015	161676	260415
<b>EF35: Familienstand</b>					
Ledig	320372	316089	314543	114775	193766
Verheiratet	402872	400271	398925	218968	243468
Verwitwet	60638	63761	60231	5760	39280
Geschieden	43675	43718	43236	25569	26561
<b>EF30: Alter (Gruppiert)</b>					
unter 20	170680	169225	169834	12875	102312
20-25	48516	46355	46354	29425	28886
25-30	45048	43899	43698	32309	26619
30-35	60139	59635	59256	47450	35948
35-40	69400	69026	68699	55687	41789
40-45	65537	65153	64879	53230	39698
45-50	58921	58566	58312	47123	35717
50-55	55294	54842	54627	40868	33447
55-60	47431	47009	46771	27912	28740
60-65	61864	61437	61087	14039	37863
65-70	47544	47416	47043	2632	29272
70-75	44741	44912	44340	1125	27800
76 und älter	52441	56365	52037	397	34984

<sup>10</sup> gewichtet mit Haushaltsgewicht (compute hhgew=ef751\*100/70; weight by hhgew)<sup>11</sup> gewichtet mit Personengewicht (compute persgew=ef750\*100/70; weight by persgew)

<b>EF44: Staatsangehörigkeit</b>					
Bosnien-Herzegowina	1842	1884	1830	841	873
Frankreich	1345	1381	1321	721	688
Griechenland	3812	3830	3778	2002	1766
Grossbritannien	1153	1224	1131	763	565
GUS	3679	3845	3661	1335	1947
Italien	7276	7424	7233	3824	3516
Kroatien	3061	3083	3018	1731	1462
Niederlande	1118	1195	1085	595	518
Österreich	1829	1933	1790	1117	898
Polen	3060	3221	3023	1638	1592
Portugal	1215	1251	1210	699	568
Rumänien	885	917	869	516	494
Übrige EU <sup>12</sup>	938	974	909	518	471
Serbien und Monteneg.	4364	4496	4341	1773	2097
Slowakei etc. <sup>13</sup>	926	967	910	569	469
Spanien	1271	1315	1266	697	604
Türkei	22543	22947	22491	7815	10347
Übriges Europa <sup>14</sup>	3296	3393	3252	1345	1643
Marokko	678	711	675	251	334
Sonstiges Afrika	1527	1660	1513	624	747
USA	1137	1177	1101	601	606
Sonstiges Amerika	1038	1101	1022	470	536
Iran	920	938	913	406	458
Sonst. Naher Osten <sup>15</sup>	2059	2134	2044	493	938
Vietnam	936	954	930	409	426
Sonst. Südasien <sup>16</sup>	1980	2103	1966	664	963
Ostasien <sup>17</sup>	1311	1392	1296	520	649
Übrige Welt, Staatenlos	1085	1124	1066	441	521
Deutscher <sup>18</sup>	751273	745265	741293	331696	466379
<b>EF338: Überw. Lebensunterhalt</b>					
Erwerbs-/Berufst.	336867	332124	331262	332124	203085
Arbeitslosengeld/-Hilfe	31208	31164	31068	1692	18469
Rente, Pension	183250	185963	181595	5110	114780
Unterhalt	248959	245244	246547	22325	149202
Eigenes Vermögen	3769	3692	3624	443	2333
Sozialhilfe	13833	15658	13799	1204	8760
Leist. Pflegevers.	588	1320	581	67	815
Sonst. Unterstützung	9082	8677	8459	2108	5631

<sup>12</sup> Belgien, Dänemark, Finnland, Irland, Luxemburg, Schweden<sup>13</sup> Tschechische Republik, Ungarn<sup>14</sup> Schweiz, Sonstiges Ost- und Mitteleuropa, Sonstiges Westeuropa<sup>15</sup> z.B. Libanon, Jordanien, Syrien, Irak, Israel<sup>16</sup> z.B. Afghanistan, Indien, Kambodscha, Laos, Pakistan, Thailand, Sri Lanka<sup>17</sup> z.B. China, Hongkong, Indonesien, Japan, Korea, Macao, Philippinen<sup>18</sup> ohne weitere Staatsbürgerschaft

<b>EF504: Erwerbstyp</b>					
Erwerbstätig	370018	365073	363804	365073	223478
Erwerbslose (EU) <sup>19</sup>	34989	34920	34799	0	20639
Sonst. Erwerbslose	5874	5839	5829	0	3461
Nichterwerbspersonen	416675	418009	412504	0	255497
<b>EF127: Stellung im Beruf</b>					
Selbst. ohne Beschäftigten.	18797	18384	18347	18384	11374
Selbst. mit Beschäftigten	18549	18205	18116	18205	11231
Mithelf. Familienangehöriger	4227	4149	4158	4149	2568
Beamter, Richter	19702	19105	19139	19105	12124
Angestellter	181422	177643	177978	177643	110237
Heim-(Arbeiter)	107377	107792	106641	107792	64048
Kaufm.-techn. Auszubild.	8294	7931	7989	7931	5021
Gewerbl. Auszubildender	7821	7727	7684	7727	4754
Zeit-/Berufssoldat	2005	2184	1953	2184	1070
Grundwehrdienstleist.	824	980	823	980	440
Zivildienstleistende	1000	973	976	973	611
Entfällt	457539	458768	453131	0	279597
<b>EF372: Höhe d. Nettoeinkommens im April</b>					
Unter 150 Euro	20150	20375	19952	4660	12367
150 bis unter 300 Euro	35332	35114	34931	14096	21327
300 bis unter 500 Euro	55938	55183	54795	26529	33975
500 bis unter 700 Euro	63700	62636	62474	23628	38463
700 bis unter 900 Euro	59834	59334	59160	24629	36212
900 bis unter 1100 Euro	67999	67621	67313	35111	41224
1100 bis unter 1300 Euro	69005	68740	68251	41578	41889
1300 bis unter 1500 Euro	57072	56832	56398	38472	34620
1500 bis unter 1700 Euro	43106	42966	42490	31816	26265
1700 bis unter 2000 Euro	38012	37955	37415	29319	23299
2000 bis unter 2300 Euro	27255	27502	26760	21547	16923
2300 bis unter 2600 Euro	19713	19931	19268	15899	12385
2600 bis unter 2900 Euro	10003	10103	9784	8098	6295
2900 bis unter 3200 Euro	9350	9241	9085	7540	5793
3200 bis unter 3600 Euro	7208	7131	7008	6073	4468
3600 bis unter 4000 Euro	4737	4684	4619	4107	2926
4000 bis unter 4500 Euro	3775	3676	3614	3227	2327
4500 bis unter 5000 Euro	2573	2499	2464	2175	1588
5000 bis unter 5500 Euro	1736	1665	1656	1492	1064
5500 bis unter 6000 Euro	1371	1306	1292	1148	838
6000 bis unter 7500 Euro	1485	1403	1391	1247	912
7500 bis unter 10000 Euro	1266	1199	1192	1082	772
10000 bis unter 18000 Euro	895	839	835	753	546
18000 Euro und mehr	576	534	529	469	354
Selbst. Landwirt	2669	2657	2651	2657	1653
Kein Einkommen	190631	188696	189906	1226	113802
Keine Angabe	32167	34017	31704	16494	20788

<sup>19</sup> sofort verfügbar (EU-Definition)

<b>EF287: Höchster allg. Schulabschluss</b>					
Haupt-/Volksschule	299780	300552	298116	126193	182628
POS	50317	49962	50045	36647	30472
Realschulabschluss	129727	128170	127939	88252	79110
Fachhochschulreife	26617	26024	25955	18948	16096
Abitur/Fachabitur	112076	107308	107111	74581	67561
Angabe fehlt	5554	5570	5413	2951	3399
Entfällt	203486	206254	202356	17501	123809
<b>EF289: Höchster berufl. Ausbildungsabschluss</b>					
Anlernausbildung	10829	10841	10721	4549	6594
Berufsvorbereitungsjahr	1345	1357	1339	649	823
Lehrausbildung	306094	304363	303378	178130	186662
Berufsfachschule	20338	20125	20057	12860	12426
Meister, Techniker	39260	38882	38702	27425	24028
Fachschule chem. DDR	10337	10225	10253	6203	6263
Verwaltungsfachhochschule	5275	5158	5141	3890	3223
Fachhochschule	26855	26245	26109	19309	16314
Hochschulabschluss	37244	36057	35930	27797	22484
Promotion	6029	5753	5705	4464	3673
Angabe fehlt	10940	10826	10728	6871	6676
Entfällt	353012	354009	348873	72926	213909
<b>EF138: Vollzeit-/Teilzeittätigkeit</b>					
Vollzeittätigkeit	292310	288895	287308	288895	176360
Teilzeittätigkeit	77708	76177	76495	76177	47118
Entfällt	457539	458768	453131	0	279597



**Anhang B: SPSS-Syntax für die Plausibilitätsprüfungen des Scientific Use Files 2002**

\* Plausibilitätsprüfung: Vergleich mit den Fachserien des Statistischen Bundesamtes MZ2002.

```
get file = ['SPSS-Datenfile'].
```

\*Bildung des Personengewichts.

```
compute persgew=ef750*(100/70).
```

\*Bildung des Haushalts- und Familiengewichts.

```
compute hhgew=ef751*(100/70).
```

```
title 'Plausibilitätsprüfung'
```

```
subtitle 'Fachserie 1, Reihe 3 (Haushalte und Familien).'
```

```
weight by hhgew.
```

```
recode ef30
```

```
(0 thru 24=1)(25 thru 34=2)(35 thru 44=3)(45 thru 54=4)
```

```
(55 thru 64=5) (65 thru 74=6) (75 thru 95=7)into ef30b.
```

\*Filter: Bevölkerung in Privathaushalten.

```
select if (ef506=1).
```

```
fre ef1 ef30b ef32 ef35 ef521 ef541 ef542.
```

\*Filter: Bevölkerung in PHH und Haushaltsbezugsperson.

```
temporary.
```

```
select if (ef506=1) and (ef507=1).
```

```
fre ef44.
```

```
weight off.
```

```
*****.
```

```
subtitle 'Fachserie 1, Reihe 4.1.1 (Stand und Entwicklung der EWT).'
```

```
weight by persgew.
```

\*Filter: Bevölkerung am Ort der Hauptwohnung.

```
temp.
```

```
select if (ef505=1 or ef505=2).
```

```
fre ef127.
```

\*Filter: Bevölkerung am Ort der Hauptwohnung, nur Erwerbslose.

```
temp.
```

```
select if (ef505=1 or ef505=2) and (ef504=2 or ef504=3).
```

```
fre ef216 ef217 ef239 ef248.
```

\*Filter: Bevölkerung am Ort der Hauptwohnung, nur Erwerbslose (Arbeitssuchend).

```
temporary.
```

```
select if (ef505=1 or ef505=2)and(ef504=2 or ef504=3)and(ef208=1).
```

```
fre ef213.
```

```

recode ef129 (11 thru 50=1)(101 thru 145=2)(151 thru 372=3)
(401 thru 454=4)(501 thru 555=5)(601 thru 642=6)
(651 thru 672=7)(701 thru 748=8)(751 thru 753=9)
(801 thru 990=10).
value label ef129 1 'Land- und Forstwirtschaft, Fischerei'
2 'Bergbau'
3 'Verarbeitendes Gewerbe'
4 'Energie- und Wasserversorgung'
5 'Handel, Gastgewerbe'
6 'Verkehr, Nachrichten'
7 'Kredit- und Versicherungsgewerbe'
8 'Grundstueck etc'
9 'Öffentliche Verwaltung'
10 'Öffentliche und private Dienstleistungen'.

*Filter: Bevoelkerung am Ort der Hauptwohnung, nur Erwerbsstaetige.
temporary.
select if (ef505=1 or ef505=2)and(ef504=1).
fre ef129.

*Filter: Bevoelkerung am Ort der Hauptwohnung,
nur Abhaengige Erwerbsstaetige mit Auszubildenden.
temporary.
select if
(ef505=1 or ef505=2)and(ef504=1)and(ef127>=4 and ef127<=8)).
fre ef138 ef139 ef287 ef289.

weight off.
*****
subtile 'Fachserie 1, Reihe 4.1.2 (Beruf, Ausbildung und Arbeitsbedingun-
gen der Erwerbsstaetigen)'.
weight by persgew.

*Filter: Bevoelkerung am Ort der Hauptwohnung und Erwerbsstaetige.
temporary.
select if ((ef505=1 or ef505=2)and(ef504=1)).
fre ef128.
weight off.

* Bildung des Personengewichts für die 0.45%-Untersichprobe.
comp persgew_u=ef755*(100/70).
weight by persgew_u.

*Filter: Bevoelkerung am Ort der Hauptwohnung und Erwerbsstaetige.
temporary.
select if ((ef505=1 or ef505=2)and(ef504=1)).
fre ef147 ef150.
weight off.

```

## **Anhang C: Weiterführende Literatur**

### **Informationen zum Mikrozensus und seiner Konzeption:**

Emmerling, Dieter/Riede, Thomas (1997): 40 Jahre Mikrozensus. *Wirtschaft und Statistik* (3):160-174.

Hartmann, P. (1989): Der Mikrozensus als Datenquelle für die Sozialwissenschaften. *ZUMA-Nachrichten* Nr. 24, Mai 1989: 6-25.

Heidenreich, H.-J.(1994): Hochrechnung des Mikrozensus ab 1990. In: Gabler, S./Hoffmeyer-Zlotnik, J./Krebs, D. (Hrsg.): *Gewichtung in der Umfragepraxis*. Westdeutscher Verlag, Opladen 1994, S. 112-123.

Lüttinger, P./Riede, T.(1997): Der Mikrozensus: amtliche Daten für die Sozialforschung, *ZUMA-Nachrichten* Nr. 41, November 1997: 19-43.

Meyer, Kurt (1994): Zum Auswahlplan des Mikrozensus ab 1990. In: Gabler, S./Hoffmeyer-Zlotnik, J./Krebs, D. (Hrsg.): *Gewichtung in der Umfragepraxis*. Westdeutscher Verlag, Opladen 1994, S. 106-111.

Statistisches Bundesamt (Hrsg.)(1997): *Arbeitsunterlagen zum Mikrozensus. Das Erhebungsprogramm des Mikrozensus seit 1957*. Wiesbaden, September 1997.

Statistisches Bundesamt (Hrsg.)(1989): *Inhaltliche Fragen bevölkerungsstatistischer Stichproben am Beispiel des Mikrozensus. Bericht zur Konferenz vom 21. und 22. Oktober 1988. Heft 10 der Schriftenreihe Ausgewählte Arbeitsunterlagen zur Bundesstatistik*. Wiesbaden 1989.

Schmidt, S. (2000): *Erwerbstätigkeit im Mikrozensus. Konzept, Definition, Umsetzung*. ZUMA-Arbeitsbericht 2000/01.

### **Zur Weitergabe faktisch anonymisierter Einzeldaten:**

Müller, W./Blien, U./Knoche, P./Wirth, H. (1991): *Die Faktische Anonymität von Mikrodaten*, (Band 19 der Schriftenreihe *Forum der Bundesstatistik*). Metzler-Poeschel, Stuttgart 1991.

Wirth, H. (1992): *Die faktische Anonymität von Mikrodaten: Ergebnisse und Konsequenzen eines Forschungsprojektes*. ZUMA-Nachrichten Nr. 30, Mai 1992.

### **Zur Fehlerrechnung im Mikrozensus:**

Krug, W./Nourney, M./Schmidt, J. (1994): *Wirtschafts- und Sozialstatistik. Gewinnung von Daten.*, München: Oldenbourg (3. völlig neu bearb. Auflage)

Statistisches Bundesamt (2001): *Bevölkerung und Erwerbstätigkeit, Fachserie 1, Reihe 4.1.1, Stand und Entwicklung der Erwerbstätigkeit 2001 (Ergebnisse des Mikrozensus)*, Stuttgart: Metzler-Poeschel.

### **Ergebnisse des Mikrozensus 2001:**

Statistisches Bundesamt (2002): *Bevölkerung und Erwerbstätigkeit, Fachserie 1, Reihe 3, Haushalte und Familien 2000 (Ergebnisse des Mikrozensus)*, Stuttgart: Metzler-Poeschel.

Statistisches Bundesamt (2002): *Bevölkerung und Erwerbstätigkeit, Fachserie 1, Reihe 4.1.1, Stand und Entwicklung der Erwerbstätigkeit 2000 (Ergebnisse des Mikrozensus)*, Stuttgart: Metzler-Poeschel.

Statistisches Bundesamt (2002): *Bevölkerung und Erwerbstätigkeit, Fachserie 1, Reihe 4.1.2, Beruf, Ausbildung und Arbeitsbedingungen der Erwerbstätigen 2000 (Ergebnisse des Mikrozensus)*, Stuttgart: Metzler-Poeschel.

**Vergleichbarkeit der Mikrozensen:**

Kollinger, N./Shahla, H. (2004): Vergleich der Mikrozensen 2000 und 2001. ZUMA-Methodenbericht Nr. 2004/09.

Lengerer, A./Hubert, T./Kollinger, N. (2005): Vergleich der Mikrozensen 2001 und 2002. ZUMA-Methodenbericht Nr. 2005/02.

Kollinger, N./Kohlmann, A. (2003): Vergleich der Mikrozensen 1998 und 2000. ZUMA-Methodenbericht Nr. 2003/08.

Lehnert, D./Kohlmann, A. (2003): Vergleich der Mikrozensen 1997 und 1998. ZUMA-Technischer Bericht Nr. 2003/02.

Schimpl-Neimanns, B./Schmidt, K. (1998) Vergleichbarkeit der faktisch anonymisierten Mikrozensen 1989, 1991, 1993, 1995. ZUMA-Technischer Bericht Nr. 98/09.

Staudenmaier B./ Granato, N./Schmidt, S. (2000): Vergleich der Mikrozensen 1995 und 1996. ZUMA-Technischer Bericht Nr.2000/08.

Staudenmaier B./Schmidt, S. (2000): Vergleich der Mikrozensen 1996 und 1997. ZUMA-Technischer Bericht Nr. 2000/09.

Darüber hinaus sind Veröffentlichungen in der Zeitschrift 'Wirtschaft und Statistik' des Statistischen Bundesamtes zu finden. Ein umfassendes Informationsangebot bietet die Abteilung Mikrodaten, ZUMA an:

<http://www.gesis.org/Dauerbeobachtung/Mikrodaten/index.htm>.

Publikationen, die auf den Scientific-Use-Files der Mikrozensen basieren finden sich auf der Webseite

[http://www.gesis.org/Dauerbeobachtung/Mikrodaten/Publikationen/MZ\\_Literatur.htm](http://www.gesis.org/Dauerbeobachtung/Mikrodaten/Publikationen/MZ_Literatur.htm).